



Rehabilitation of speech disorders following glossectomy, based on ultrasound visual illustration and feedback

Marion Girod-Roux, Thomas Hueber, Diandra Fabre, Silvain Gerber, Mélanie Canault, Nathalie Bedoin, Audrey Acher, Nicolas Béziaud, Eric Truy, Pierre Badin

► To cite this version:

Marion Girod-Roux, Thomas Hueber, Diandra Fabre, Silvain Gerber, Mélanie Canault, et al.. Rehabilitation of speech disorders following glossectomy, based on ultrasound visual illustration and feedback. *Clinical Linguistics & Phonetics*, 2020, 34 (9), pp.826-843. 10.1080/02699206.2019.1700310 . hal-01977670

HAL Id: hal-01977670

<https://hal.science/hal-01977670>

Submitted on 15 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Rehabilitation of speech disorders following glossectomy based on ultrasound visual feedback

Marion Girod-Roux^{1,2}, Thomas Hueber¹, Diandra Fabre¹, Silvain Gerber¹,
Mélanie Canault³, Nathalie Bedoin³, Audrey Acher⁴, Nicolas Béziaud²,
Eric Truy^{5,6}, Pierre Badin¹

¹ GIPSA-Lab, UMR 5216, CNRS – Grenoble Alpes Univ., France

² Centre Médical Rocheplane, Saint-Martin d'Hères, France

³ Laboratoire Dynamique du Langage, UMR 5596, CNRS, Université Lumière Lyon 2
& Institut des Sciences et Techniques de la Réadaptation, Université Claude Bernard, Lyon 1,
France

⁴ Unité Neuro-Vasculaire, Pôle Psychiatrie-Neurologie-Rééducation, CHU Grenoble Alpes,
France

⁵ Département d'ORL, de Chirurgie cervico-maxillo-faciale et d'Audiophonologie,
Groupement Hospitalier Edouard Herriot, Lyon, France

⁶ ImpAct (Integrative multisensory perception Action cognition team) Lyon Neuroscience
Research Center - CRNL (Inserm U1028, CNRS UMR5292), France

Abstract

Inaccurate positioning or modified shape of speech articulators due to intraoral surgery can induce persistent speech disorders. The literature emphasises the importance of the link between speech perception and production, and in particular of auditory and proprioceptive feedback, both for language acquisition and maintenance. Some theories postulate the existence of speech internal models – or representations – coding the relations between the articulatory movements, their proprioception and their acoustic outcome. Therefore, we hypothesise that part of articulatory disorders may be ascribed to some degradation of the efficiency of the sensory channels, especially the proprioceptive ones, or an erroneous speaker's internal representation. Since many studies have revealed that subjects are somehow able of articulatory awareness and of *tongue reading*, using the vision of articulators can be assumed to enhance sensory feedback channels and constitute a promising perspective for the remediation of articulation disorders. In the article, we describe a clinical study on the efficiency of speech therapy based on the visualization of tongue movements captured by ultrasound imaging for the rehabilitation of patients who underwent intraoral surgery subsequent to tongue cancer. We explored two rehabilitation paradigms: *visual illustration*, which consists in providing the patient with pre-recorded ultrasound image sequences showing the target tongue movement, and *visual feedback*, that is the visualization in real-time by the patient of his/her own tongue. Based on a cohort of 10 patients, we demonstrate the benefit of complementing the illustration paradigm with visual feedback.

Keywords

Speech therapy, tongue ultrasound imaging, biofeedback, articulation disorder, assistive technologies.

1 Introduction

Speech perception and production mechanisms are concurrently developed during the language acquisition process and remain closely related in its maintenance. Various studies demonstrate how critical the perception of others and of oneself is for language acquisition and maintenance (e.g. Joseph S Perkell (2012) or Turgeon, Prémont *et al.* (2015) for auditory perception, and Mills (1987) for visual perception). Joseph S. Perkell, Guenther *et al.* (2000) develop a theory of the segmental component of motor control which considers that the programming of the articulatory movements to reach auditory targets (*i.e.* phonemes or segments) relies on an internal model. This model, also called *effeient copy*, predicts the acoustic consequences of articulatory movements and adjusts the motor commands to minimize the discrepancy between the prediction and both the auditory feedback and somatosensory feedback.

The link between perception and production, that is to say the perceptuo-motor nature of speech as described by Schwartz, Basirat *et al.* (2012), is made explicit in various studies that propose internal models of speech representation. For instance, Treille, Vilain *et al.* (2016) suggest that the visual processing of audible but not visible movements (such as tongue movements) induces ‘motor and visual mental simulation of the perceived actions to facilitate their recognition’.

Numerous studies have highlighted and quantified the contribution of the vision of visible articulators (such as lips, jaw, face, tongue tip, teeth) to the perception of speech (e.g. Erber (1975), Sumbly & Pollack (1954), Benoît & Le Goff (1998); see Badin, Tarabalka *et al.* (2010) for a review). Interestingly, Badin *et al.* (2010) demonstrate that, in addition to the well documented *lip reading* ability, humans possess a *tongue reading* ability as well, *i.e.* they are able to use the vision of tongue inside the oral cavity for the recognition of phonemes in adverse (e.g. noisy) conditions.

This theoretical and experimental work has demonstrated the links between speech production and perception, as well as the ‘naïve’ capacities to exploit visual information from an inside articulator whose natural representation, almost exclusively based on proprioception, remains partial. This framework served as the starting point for several studies focusing on the interest of an articulatory visual representation that would help the rehabilitation of speech disorders. Indeed, speech disorders may be attributed to a deterioration of the efficiency of the sensory channels, especially the proprioceptive one, necessary for the acquisition or maintenance of speech production capabilities, or to a lack of internal representations of the speaker. These deficiencies may be related to deafness, motor disability, developmental disorders or orofacial surgery. Using the vision of articulators as a complement of sensory feedback may thus offer an interesting opportunity for the rehabilitation of speech disorders.

1.1 Articulatory visual illustration and feedback

In this article we distinguish two articulatory visualisation paradigms. The first one, called ‘articulatory visual illustration’, aims to provide the patient with an intuitive visualisation of a targeted lingual movement, recorded on a reference speaker (typically someone with no speech disorders such as the speech therapist). Through various collaborations between engineers, researchers and speech-language pathologists (SLPs) various software applications have emerged. These applications are based either on stylized movements such as Diadolab (Menin-Sicard, Sicard *et al.* (2016)), ‘The vocal tract in action’ (Canault, 2010)¹, or on real articulatory data such as Massaro & Light (2004), Engwall (2008), Fagel & Madany (2008), Badin *et al.* (2010) or Hueber (2013). Chen, Johnson *et al.* (2016) review the literature on the contribution

¹ This software can be tested at <http://anatomie3d.univ-lyon1.fr/webapp/website/website.html?id=3346735&pageId=223201>

of these visual illustration techniques in various populations: hearing-impaired people with cochlear implants, children with stigmatism or adults with aphasia. In each case, visual illustration seems to accelerate the learning process. It also appears to be easily assimilated by the patient who manages to associate what (s)he sees on the screen - which stands in the articulatory internal space of another speaker - with his/her own articulation.

In contrast, the second paradigm, called ‘articulatory visual feedback’², aims to provide the patient with a visual representation of his/her own tongue movements, in his/her own articulatory space and with his/her own morphology. The goal is to help him/her better understand and correct his/her gestures, in particular by providing him/her with his/her own complementary visual articulatory information, along with the practitioner's instructions. The two main techniques actually used for visual biofeedback in speech therapy are electropalatography (EPG)³ and ultrasound imaging (US).

EPG has been used for three decades as biofeedback in speech therapy (*see* Gibbon (2013) for an exhaustive review). All studies conclude to some positive effects of EPG, though most of them do not involve control subjects and statistical evaluation. Drawbacks of EPG are the need to build a customized palate for each patient as well as the invasiveness of the palate and connecting wires.

Ultrasound imaging is a very interesting technique to observe tongue movements during speech production in real time, as it is harmless and minimally invasive for speakers (*e.g.* Epstein (2005) and recently Jonathan L. Preston, Holliman-Lopez *et al.* (2018)). The reader is referred to Stone (2005) for a complete description of the use of ultrasound imaging in phonetic research, experimental setup, analysis techniques, etc. More than three decades ago, Shawker & Sonies (1985) suggested to use ultrasound images of the tongue for the rehabilitation of English /ɪ/ for a nine year old patient. Since this first case study, English /ɪ/ rehabilitation with ultrasound remains a widely studied topic (Adler-Bock, Bernhardt *et al.* (2007), Jonathan L Preston & Leaman (2014), Byun, Hitchcock *et al.* (2014), Cavin (2015), Jonathan L Preston, McAllister *et al.* (2018)). Modha, Bernhardt *et al.* (2008), Jonathan L Preston, Brick *et al.* (2013), Joanne Cleland, James M. Scobbie *et al.* (2015), Jonathan L Preston, Maas *et al.* (2016) or Cleland, Scobbie *et al.* (2017) studied the use of ultrasound in speech sound disorders or apraxia of speech. Studies have also been conducted in cases such as severe hearing impairment (Adler-Bock *et al.* (2007), Bernhardt, Gick *et al.* (2003) or Gallagher (2013)), Down syndrome (Fawcett, Bacsfalvi *et al.* (2008)), cleft palate (Roxburgh, Cleland *et al.* (2016)), acquired apraxia of speech post stroke (Jonathan L Preston & Leaman (2014), Acher, Fabre *et al.* (2016) and Haldin, Acher *et al.* (2017)) or partial glossectomy (Blyth, McCabe *et al.* (2016)).

In summary, the literature on ultrasound visual feedback reflects a great diversity of experimental designs. Thus, depending on speech disorders and studies, various protocols regarding number, duration and/or content of sessions, or number and age of participants, have been set up. Importantly, most studies have been conducted in English-speaking countries. Moreover, the performance was mostly evaluated through speech assessments although acoustic (audio recordings) or articulatory (contours segmentation of images) analyses sometimes supplement these reports (Joanne Cleland, James M Scobbie *et al.* (2015)). Besides,

² As the feedback or illustration used in the present study are of an articulatory nature only, the term « articulatory » will not be specified in the rest of the article.

³ Electromagnetic articulography (EMA) has also been used by Katz & McNeil (2010) and Mental, Carey *et al.* (2017) but remains too invasive and not convenient enough to be concretely used in a clinical environment.

the experimental designs differed from one study to another: while Bernhardt, Bacsfalvi *et al.* (2008) alternated traditional rehabilitation and ultrasound visual feedback, Roxburgh *et al.* (2016) proposed a comparative study between the ultrasound visual feedback and the illustration by an articulatory talking head, whereas Cleland, McCron *et al.* (2013) and Bacsfalvi, Bernhardt *et al.* (2007) aimed to compare two main visual feedback devices, EPG and ultrasound. As a matter of fact, the recent study by Jonathan L Preston *et al.* (2018), aiming to compare two types of ultrasound biofeedback, is the only one that led to a statistical analysis of the outcomes. Finally, Blyth *et al.* (2016) led the only study of ultrasound biofeedback for two glossectomized patients.

Despite their great diversity, all these studies converge towards the same conclusion: the use of ultrasound seems to have a therapeutic benefit in speech rehabilitation as it enhances the positive impact of speech therapy, especially in the early stages (*e.g.* Blyth *et al.* (2016), Jonathan L Preston *et al.* (2018)). This contribution, yet, remains subtle and difficult to measure. It is therefore necessary to conduct larger studies focused on specific disorders to hope to gather convincing arguments.

This is the main motivation of the present study which focusses on speech disorders occurring after oral surgery.

1.2 Speech disorders related to oral surgery with partial glossectomy

When a cancerous lesion appears on the tongue, it can be removed by a surgical procedure called partial glossectomy, which can be extended to the mouth floor and/or the mandible. If the resected specimen is too large and the functional outcomes are expected to be too severe, the surgeon may be required to replace the resected specimen by a flap made from a patient's piece of muscle, bone and/or skin. These surgical gestures are often supplemented by chemoradiotherapy (CRT) which severely alters the remaining tissues and the patients' quality of life (QoL). Because of the intraoral resection and the CRT adjuvant treatment, patients may often experience difficulties in both eating (since tongue is involved in oral preparation, propulsion and swallowing of the bolus) and communicating (as tongue is deeply involved in speech production). The vocal tract has been modified in terms of shape and volume, strength and sensibility of the remaining tissues, as well as in amplitude and velocity of their movements. Subsequently, speech is often affected by a loss of tone, precision and speed, which can alter the production of certain phonemes as described by Perrier, Savariaux *et al.* (1999). Acher, Perrier *et al.* (2014) report possible alterations in French patients after partial glossectomy. Thus, consonants involving tongue movements in the point and the manner of articulation, *i.e.* alveolar, post-alveolar and velar consonants, are naturally affected by these anatomical changes. For instance, a significant noise can be noticeable during the holding phase of these obstruent consonants. Speech therapy is then necessary to help the patient to accurately produce speech and master his/her new articulatory space.

1.3 Goal of the present study

The purpose of the present study was thus to evaluate - in a clinical environment - the efficiency of visual tongue illustration and/or feedback in speech rehabilitation after intra-oral surgery including partial glossectomy. For this purpose, ten patients who have undergone intra-oral surgery have been included in this study over a period of two years (2016 - 2018). They followed speech therapy sessions based either on visual illustration or on a combination of visual illustration and feedback, as detailed in Section 2.1, and were regularly assessed.

2 Materials and methods

This study was carried out at the Rocheplane rehabilitation centre in Saint Martin d'Hères, France, with the agreement of the ethical committee Lyon Sud Est II (69HCL15_0736). This rehabilitation centre receives patients when they are discharged from the hospital after their surgery and provides them with intensive speech and swallowing therapy.

Three SLPs were involved in this project. Two of them (MGR and BG) were alternatively responsible for the patients' rehabilitation, according to their schedules, whereas two of them (MGR and OC) were in charge of monitoring the patients' progress.

2.1 Therapy sessions

This study aimed to compare the progress achieved with two speech rehabilitation protocols.

Two protocols were followed alternatively. In the protocol providing *visual illustration only* (*IL protocol*), the patient visualized the target/gesture produced by a reference speaker but relied only on his/her acoustic and somatosensory feedback to understand why his/her gesture was wrong and how (s)he should correct it. In the protocol providing both visual illustration and feedback (*FB+I protocol*), the addition of a real time visual feedback of his/her own gesture is expected to fasten the recalibration process.

Thirty minutes therapy sessions were conducted once or twice a day, weekend excluded. The patients were divided in two cohorts (FI and IF). The patients of the FI cohort followed a first series (S1) of 10 therapy sessions with the *FB+I* protocol, and then a second series (S2) of 10 sessions with the *IL* protocol. The order of the series in the IF cohort was reversed. A speech assessment was performed by the SLP at baseline, prior to the first series (at time T0) and then at the end of each series S1 and S2 (times T1 and T2). Speech therapy sessions were usually conducted at a mean frequency of about 5 sessions per week (2 to 10 depending on patient overall condition, but constant for each patient, *cf.* Table 1). Patients were assigned to the FI or IF cohorts using stratified random sampling based on their T0 assessment scores in order to obtain similar means for both cohorts.

2.2 Patients

Twelve adult patients were initially included in the study: eight males and four females. One male left before the end of the study whereas one female experienced problems of concentration during the T1 assessment and did not read the right words; their scores are thus not reported nor analysed in this study. Finally, ten patients (seven males and three females) were retained (average age of 59.4 years for the FI cohort and 58.6 years for the IF cohort). Details of surgery and treatment for each patient are presented in Table 1. The patients included in this study did not report any speech disorder or speech therapy prior to surgery. Note that all of them were explained the protocol and signed an informed consent before being included in the study, as described in the protocol allowed by the ethical committee.

Table 1. General information on patients and their cohorts. IF- refers to the cohort IL / FB+I, while FI- refers to FB+I / IL.

Code patient	IF-001	FI-002	FI-003	IF-004	IF-005	FI-007	FI-008	IF-009	FI-010	IF-012
Age (years)	47	54	52	58	58	53	75	70	63	60
Gender	M	M	F	F	M	M	M	M	F	M
Inclusion day after surgery	J+20	J+21	J+14	J+28	J+24	J+26	J+18	J+13	J+23	J+31
Type of cancer	T4 Floor of mouth + mandible	T2 Junction base / body of tongue	T1 Floor of mouth	T3 Lateral border of the tongue	T4 Anterior left floor of mouth	T4 Right oropharynx	T3 Amygdalotomical sulcus + base of tongue	T3 Lateral border + hemibase of tongue + floor of mouth	T2 Floor of mouth / Tongue anterior right sulcus	T4 Floor of mouth
Surgery	FOM + Partial glossectomy + Mandibulectomy (transoral + transcervical)	Oropharyngectomy + Partial glossectomy (transcervical + transoral)	FOM excision	Lateral glossectomy	FOM excision + anterior glossectomy + mandibulectomy	Oral cavity / Oropharyngeal resection / Mandibulectomy + BOT resection (transoral + transcervical)	Oral cavity / Oropharyngeal resection / Mandibulectomy (transoral + transcervical)	Lateral oropharyngectomy + Partial glossectomy + BOT partial resection	Glossectomy + anterior FOM resection	FOM
	Flap	Flap	Local Flap	Free Flap	Flap (AB)	Flap (GP)	Flap (GP)	Local FAMM Flap	Local flap	Tongue
Chemoradiation (CRT)	CRT during study	CRT during study	Not planned	CRT during study	CRT during study	CRT during study	RT during study	Not planned	Not planned	CRT planned after the study
Dental status	Toothless on the inferior left arch	Toothless (superior arch)	Toothless	Complete dental appliance	Toothless on the inferior left arch	Toothless	Toothless	Toothless	Healthy and complete	Toothless
Frequency of sessions / week	5-6	6-7	6-8	2-4	4-6	4-6	3-6	6-7	10	3-7

2.3 Experimental setup

2.3.1 Visual illustration software

The *Ultraspeech-player* software (Hueber (2013)) was used in this study for visual tongue illustration. This software is dedicated to the visualization of high-speed ultrasound and video sequences of the tongue and lips, recorded synchronously with the speech audio signal, using the *Ultraspeech* software (Hueber, Chollet *et al.* (2008)). The software aims to display natural tongue movements acquired on a reference speaker, for different kinds of sequences (isolated vowels, VCV, swallowing, etc.), as illustrated in Figure 1 (top). *Ultraspeech-player* embeds a mechanism of real-time processing of audio and video streams, allowing the user to control the speed of the illustrated articulatory gesture, and that of the associated sound signal, in order to better observe it. The use of articulatory data acquired on a real speaker allows to restore realistic coarticulatory patterns. This software is free to download at <http://www.ultraspeech.com>.

For the present study, a new database was recorded and added to *Ultraspeech-player*⁴. One of the SLTs involved in the study was instructed to pronounce all the training sequences used

⁴ This database will be made publicly available if the paper is accepted.

during the therapy (see 2.4 for more details). Her tongue movements were recorded at 60fps using the Telemed Echoblaster system configured with a typical parameter set for tongue imaging (e.g. 7cm depth, with all image post-processing techniques such as frame averaging or speckle reduction disabled), with a 3-5MHz microconvex probe, maintained fixed with respect to the speaker's head (and below the chin) thanks to a slightly modified version of the helmet developed by *Articulate Instrument*⁵. Lip movements were also recorded (but not used in the therapy session) at 60fps (640×480 pixels) using an industrial CMOS camera (ImagingSource DFM22BUC03ML). The acoustic speech signal was acquired at 44.1kHz / 32bits with the low-latency (ASIO compatible) RME Fireface soundcard and a cardioid AKG C-1000S microphone. All recordings were performed in a semi-anechoic room using the *Ultraspeech 1.3* recording software.

One possible issue with ultrasound tongue images is the difficulty for the patient to make sense of them. Actually, the tongue is displayed out of any spatial context since the palate, teeth and pharynx are not visible in ultrasound images. In order to make the image more intuitive, the ultrasound image was supplemented with a schematic contour of the oral cavity (in red on Figure 1). This contour was extracted from a static MRI scan available for a female speaker. A set of transformations (rotation, shift, scaling) was then manually estimated to register the ultrasound image with the MRI contours. Using a 'trial and error' approach, the coarse-grained location of tongue-palate contacts for alveolar and velar stops was empirically checked, as well as the position of the tongue for some front and back vowels (note that the goal was not to register perfectly the two modalities together, which is a research problem on its own, but simply to make the ultrasound tongue image easier to interpret).

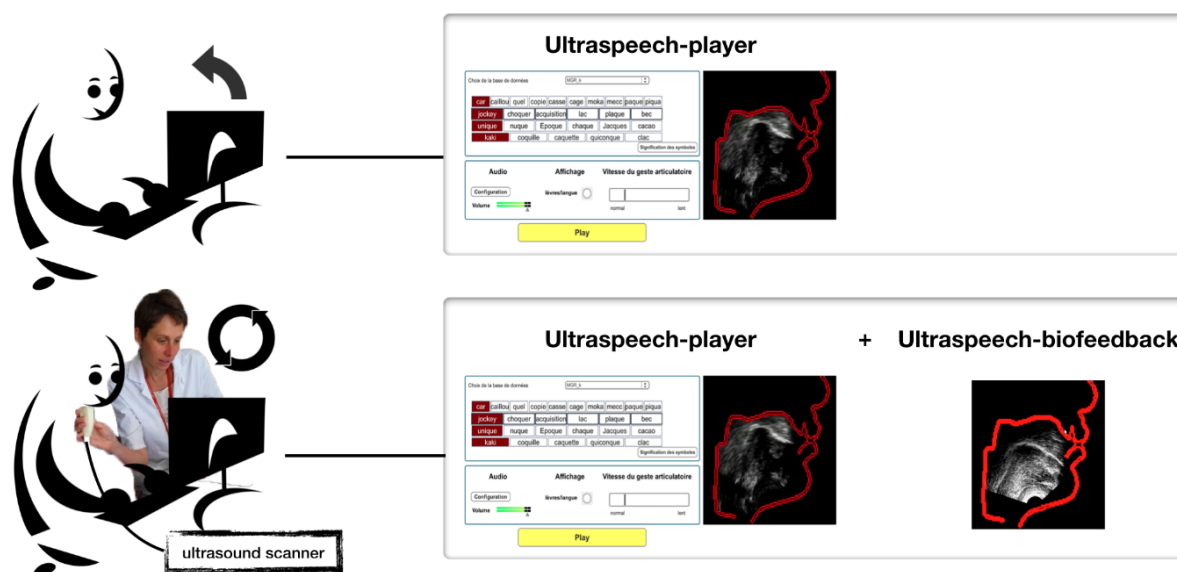


Figure 1. Top: illustration of the control interface and midsagittal ultrasound image display with sketchy vocal tract outline provided by *Ultraspeech-player* and used in the *IL* protocol. Bottom: illustration of the double interface, including the real time display of the speaker's tongue by *Ultraspeech-biofeedback*, used in the *FB+I* protocol.

⁵ <http://www.articulateinstruments.com/>

2.3.2 Visual feedback hardware and software

The same ultrasound system was used during the therapy sessions as a biofeedback tool. However, the probe stabilization helmet was not used since it may have been too uncomfortable in case of post-surgery facial oedemas. The probe was thus held manually by the SLT, as patients had a partial loss of sensitivity in the region below the chin and tongue, and were not able to maintain the ultrasound probe in the adequate position.

In order to ensure a consistent display between illustration and feedback, a new software was developed, called *Ultraspeech-biofeedback*⁶. It allows to superimpose in real-time the same schematic contour delimiting the oral cavity mentioned in section 2.3.2 on the ultrasound image (see Figure 1, bottom). Just before the first session for each new patient, the SLT took a minute to adjust the display to the intraoral morphology of the patient using the trial-and-error procedure mentioned above and stored the settings in a preset file which was reloaded at each following session and possibly updated.

2.4 Rehabilitation material

For the first half of the therapy sessions, the patients were instructed to perform Oral Motor Exercises (OME) including lips, mouth, cheeks and tongue range of movements. For the second part of the sessions, articulation of selected phonemes involving tongue (/t/, /k/, /kt/, /kl/, /st/, /sk/, /tʃ/ and /tl/) was trained using previously recorded words and sentences as described in Section 2.3.1.

2.5 Speech assessment tools

Speech assessment was carried out to evaluate patients' progress by means of tests extracted from the French BECD test battery (Auzou & Rolland-Monnoury (2006)). Inspired by the Frenchay Dysarthria Assessment (Enderby (1983)), it assesses the degree of dysarthria with different approaches. In the present study, three tests addressing isolated phonemes and words / sentences / conversation were used:

- 'Phonetic Intelligibility test' (*TPD*): the patient is instructed to read from cards a list of 52 words randomly chosen from 13 groups of four words. In each group, one word differs from the three others by one or two phonetic features (for instance 'début' [deby], 'débute' [debyt], 'des bouts' [debu], 'déboute' [debut]). The SLP selects the closest word to what (s)he hears (without seeing it) from a list of four words. In each of the 13 series, the phonetic feature errors are scored on between 0 and 8 (since two features may simultaneously be erroneous for each of the four words, e.g. [deby] vs. [debut] in the previous example). Note that the three random lists (for T0, T1 and T2) were the same for all patients.
- 'Intelligibility test' (*Intell*): it is made of three parts, based on a set of words, a set of sentences and a spontaneous conversation. The patient is instructed to read twelve words and twelve sentences randomly picked in a pile of cards. The SLP reports what (s)he hears, without seeing it, and compares it with the expected content. Finally, the patient has a short conversation with the SLP. A global score between 0 and 8 is attributed by the SLP for each part.
- 'Phonetic Analysis test' (*PhAn*): the patient is instructed to repeat a list of 33 isolated phonemes, 37 words including simple phonemes and 25 words with complex phonemes

⁶ This software will be made publicly available if the paper is accepted.

pronounced by the SLP. The SLP scores the altered phonemes for each production (33 isolated, 88 simple and 30 complex phonemes), leading to a total of 151 scores on 0-1.

The rehabilitation program is conducted by two SLPs who alternatively educate the patients. Initially, all assessments were performed at T0, T1 and T2 by only one of them (MGR) for the sake of evaluation homogeneity. However, as this evaluation was not blinded, a blind SLP (OC) was asked to rate all the patients again, based on the audio records, with the same protocol.

3 Results

In order to evaluate statistically the results of each test, it is needed to model the relation between the response variable, *i.e.* the *score* obtained by the patient, and several explanatory variables as well as their interactions, *i.e.* *cohort* (qualitative variable, 2 categories, FI and IF) and *time* (qualitative variable, 3 categories, T0, T1 and T2).

The *score* variables are integer values between 0 and 8 for *TPI* and between 0 and 8 for *Intell*, and can thus be considered as ordered categorical variables. Besides, as each patient was tested three times, the *patient* variable (10 participants) is considered a random effect for the test with repeated data. We have therefore employed *ordinal regression with random effects* for the *TPI* and *Intell* tests, using the *clmm* procedure of the *R package* (R Development Core Team (2008)) based on Tutz & Hennevoel (1996).

Then, it was checked whether the models adjusted the data correctly, *i.e.* if the empirical probabilities (*i.e.* relative frequency) were close to the probabilities estimated by the model and if they were within the range of 95% prediction. Finally, four contrasts have been estimated by means of the *lsmeans* procedure of the *R package*. For each test, the following comparative contrasts have been evaluated (note that 'FI or IF(Ti-Tj)' denotes the difference in scores in the FI or IF cohort between times Ti and Tj):

FI_IF_S1 = FI(T0-T1) *vs.* IF(T0-T1): difference in progress achieved after the first series S1, *i.e.* difference between FI cohort (*FB+I* sessions) and IF cohort (*IL* sessions) after the first series;

FI_IF_S2 = FI(T1-T2) *vs.* IF(T1-T2): difference in progress achieved after the second series S2, *i.e.* difference between FI cohort (*FB+I* protocolsessions) and IF cohort (*IL* sessions) after the second series S2;

FI_IF = $\frac{1}{2}$ [FI(T0-T1) + IF(T1-T2)] *vs.* $\frac{1}{2}$ [FI(T1-T2) + IF(T0-T1)]: difference in progress between *FB+I* protocol and *IL* sessions averaged over series S1 and S2.

For test *PhAn*, the *score* variable is the sum of the scores obtained for each of the 151 assessed phonemes, divided by 151. This, consequently, takes a nearly continuous value between 0 and 1. The *patient* variable is again considered as a random effect. For this type of scores, *beta regression with random effects* has been employed using the *glmmadmb* procedure of the *R package glmmADMB*. The four contrasts were estimated with the *glht* procedure of the *R multcomp* package based on Hothorn, Bretz *et al.* (2008). The whole analysis has been performed for both raters (MGR and OC) separately.

The analysis was first applied, for the three tests, to the scores of the ten patients by the two raters: no difference was found significant at $p=0.05$. This means that it is impossible to claim that any patient significantly benefited better from the *FB+I* sessions than from the *IL* sessions.

A cohort analysis was then performed for each cohort of patients (FI and IF), for each test (*TPI*, *Intell*, *PhAn*) and each SLP (MGR, OC). The first result is that none of the differences were significant for the *Intell* test. This was not unexpected for two reasons: as the number of words in the list was small, the SLP tended to un-volitionally learn them and eventually to answer in

a forced choice way; moreover, each score level grouped several possible errors, which reduced the chance to observe small progress. No detailed results are thus reported for this test which appeared to be less relevant as expected. The results for the other two tests (*TPI* and *PhAn*) are reported in Table 2.

The first important observation is that, for the *TPI*, the *FB+I* sessions are significantly more efficient than the *IL* sessions, on average over both S1 and S2 series, for both raters. This is a positive indication that patients' overall progress tends to be better with the *FB+I* sessions than with the *IL* sessions. A similar observation can be made from the results of the S2 series, though the difference related to rater MGR has a p-value slightly above 0.05. More surprising are the results for the S1 series, where only rater OC finds that the *IL* sessions were significantly more efficient than *FB+I* sessions, though the difference remains small. Finally, Table 2 shows that the *PhAn* test does not lead to any significant differences, except for rater MGR in session S1.

Table 2. Cohort score progress averages and standard deviations (SD); p-values of their comparisons (significant differences are marked in bold). The 'S1 + S2' columns correspond to the merging of the *FB+I* scores and of the *IL* scores from S1 and S2.

Rater	Test	S1						S2						S1 + S2					
		Cohort FI			Cohort IF			Cohort IF			Cohort FI								
		<i>FI+I</i>			<i>IL</i>		p-val	<i>FI+I</i>			<i>IL</i>		p-val	<i>FI+I</i>			<i>IL</i>		p-val
MGR	<i>TPI</i>	4.4	SD 2.7		0.8	SD 3.9	0.1154	5.0	SD 2.3		0.0	SD 2.2	0.0578	4.7	SD 2.4		0.4	SD 3.0	0.0437
OC	<i>TPI</i>	0.4	SD 2.8		0.8	SD 2.6	0.0001	4.6	SD 3.6		0.8	SD 1.3	0.0001	2.5	SD 3.7		0.8	SD 1.9	0.0001
MGR	<i>PhAn</i>	7.4	SD 5.1		3.2	SD 1.9	0.0372	2.4	SD 2.1		2.4	SD 3.2	0.8638	4.9	SD 4.5		2.8	SD 2.5	0.7602
OC	<i>PhAn</i>	2.6	SD 14.3		9.8	SD 11.6	0.4330	-4.8	SD 10.0		6.4	SD 12.8	0.4140	-1.1	SD 12.3		8.1	SD 11.6	0.3710

4 Discussion, conclusion and perspectives

Our literature review showed the importance of multiple sensory feedback channels in motor actions such as speech, including speech in glossectomized speakers. These multiple channels help maintain or correct internal models of coherent speech production and perception. We presented two paradigms of articulatory visual representation (for the rehabilitation of articulatory disorders) that were expected to improve the patients' speech ability by providing them with additional information on the defective phonemes: 'illustration' and 'articulatory visual feedback'. Based on the 10 glossectomized patients included in the present study, taking care to counterbalance the order and treatment conditions to avoid possible bias, the results of this research support the idea that the use of articulatory visual feedback significantly improves speech outcomes.

Baseline (T0) and tests (T1 and T2) have only assessed non-treated words (which included treated *and* non-treated phonemes) in order to observe generalization in conversational speech. The present study tends to show that speech was significantly improved in both cohorts with the use of US biofeedback. Indeed, in the early stage of the rehabilitation (2-3 first weeks) and especially within the S2 series, both cohorts show better speech outcomes ($p < 0.05$) at the *TPI* test with the combination of US visual feedback and illustration, than with the illustration only.

Regarding S1 series and according to rater OC, results seem to be in favour of the illustration only. At this point, we have not found any interpretation. One hypothesis is that this rater was less accustomed to glossectomized speech than the other rater MGR.

Intra-patient analysis has not been conducted as only cohort analysis led to statistically significant results, and as inter-rater agreement was low. The analysis of the *TPI*-test shows that all patients improved their speech during the study (from T0 to T2), except patient FI-008; this patient had been particularly more affected by chemoradiation, leading to a lot more pain, fatigue and lack of motivation, than the other patients. Moreover, according to both raters, the speech quality decreased for one patient (patient IF-004) within the first series (*IL*); this can be

explained by the occurrence of intercurrent medical events (in this case, a flap alteration). Furthermore, a qualitative analysis of the errors in the *TPI* shows that errors on aperture for vowels and place of articulation for consonants benefited from the use of the US visual feedback. Lastly, previous single-case studies of biofeedback intervention, including the latest of Jonathan L Preston *et al.* (2018) reported the existence of non-responders, which could be the case of patient FI-007: according to rater OC, this patient produced more errors at T1 assessment after the US visual feedback sessions.

Finally, the patients of this study have informally reported benefits – at least for the first sessions – of the US real-time feedback: they felt that the feedback helped them better imagine their tongue position, movement or shape – of whose they were not aware due to their naïve condition and their lack of tongue sensitivity following surgery. During the therapy sessions, the SLPs hold the probe under the patients' chin. Indeed, participants were unable to adequately maintain the probe due to the lack of sensitivity in the chin's area. This is consistent with Jonathan L. Preston *et al.* (2018)'s study of the undesired effects in ultrasound speech therapy, where 21% of the responses were related to the difficulties in positioning the probe, which was found uncomfortable or annoying. In agreement with Roxburgh, Scobbie *et al.* (2015), both SLPs also reported that the US biofeedback enabled them to provide positive reinforcement to their patient on their skills.

Note that, since the 10 subjects of the study were patients following intensive therapy and complementary treatment after surgery (chemotherapy and/or radiation), the frequency of the therapy sessions had to be modulated and adapted to the participants (overall condition, medical complication, pain, fatigue, medical consultation appointment, chemoradiation sessions, public holidays, etc.). Thus, even if they all have followed the study design (10 sessions of each protocol I and F), they did not have the same number of sessions per week (mean = 5 sessions / week, range = 2–10 sessions / week): this modulation could have had an impact on the results. Moreover, the heterogeneity of the cohort in terms of tumour stage, type of the surgery, need for chemoradiation, number of remaining teeth, etc., may also have a potential influence on the results, for instance by increasing the dispersion of the scores.

Another possible bias may arise from the heterogeneity of the therapist contribution to the rehabilitation and assessment tasks. Indeed, two SLPs were in charge of the patients' rehabilitation with or without US visual feedback, depending on their work load and schedule, though only one of them was in charge of the assessment, for the sake of homogeneity. A third blind-SLP was in charge of the reassessment of the 10 patients based on the audio records, so that the inter-rater reliability could be evaluated. The results show various degrees of agreement between the two raters, consistent with the findings of Roxburgh *et al.* (2016) who reported a statistically weak agreement between the scores of 22 phonetically-trained listeners for the same patients.

Regarding the assessment material, Woisard, Espesser *et al.* (2013) showed that the standard BECD test battery (Auzou & Rolland-Monnoury (2006)) suffers from limitation, in particular due to its use of real words: the raters are more likely to restore distorted sequences for real words, which may induce them to under- or over-estimate the errors, depending on phonemes. In the future, it could be more efficient to use the new intelligibility test proposed by Ghio, Lalain *et al.* (2018), based on the pronunciation of pseudo-words that neutralizes unwanted lexical and learning effects of items by the raters. However, such a test would not allow therapists to assess speech in its functional dimension of communication.

To conclude, the present study is, to the best of our knowledge, the first cohort study demonstrating the effectiveness of ultrasound visual feedback jointly used with illustration for rehabilitating glossectomized patients.

The generalisability of these results is however subject to certain limitations. For instance, the sample size remains limited and should be increased in future studies (a multicentre trial could be an efficient way to recruit more patients while maintaining strict inclusion criteria). The speech assessment method could also be improved, for instance by relying on intelligibility tests recently proposed in the literature (*e.g.* Ghio *et al.* (2018)); it could also involve a more extended set of raters (though Roxburgh *et al.* (2016) have shown that the agreement remains low). In addition, future work could also exploit recent technological development aiming at improving the displayed ultrasound tongue image. Indeed, such image may remain difficult to interpret for some patient since the tongue is displayed out of any spatial context (the palate, teeth and pharynx are not visible in an ultrasound image of the vocal tract). Fabre, Hueber *et al.* (2017), have recently proposed a machine learning- based method for animating automatically and in real-time the 3D tongue model of an articulatory talking head from ultrasound images. Validating this technique in a clinical context is part of our roadmap.

Acknowledgements

This work was funded by the Rhône-Alpes Region ARC6 program ‘Information and Communication Technologies and Innovative IT Practices’, as part of Diandra Fabre's doctoral thesis. Authors are very indebted to SLP Bérengère Gal who shared the rehabilitation work and to SLP Ondine Champavère who performed the second assessment based on audio recordings, as well as to Dr Thomas Dell’Accio who participated in the follow-up examinations.

Statement of interest

The authors report no declarations of interest.

References

- Acher, A., Fabre, D., Hueber, T., Badin, P., Detante, O., Cousin, E., Pichat, C., Loevenbruck, H., Haldin, C. & Baciú, M. (2016). Retour visuel en aphasiologie : résultats comportementaux, acoustiques et en neuroimagerie. In *XVI^{èmes} Rencontres Internationales d’Orthophonie. Orthophonie et technologies innovantes (UNADREO)* (N. Joyeux & S. Topouzkhani, editors), pp. 227-260. Paris, France: Ortho Edition.
- Acher, A., Perrier, P., Savariaux, C. & Fougeron, C. (2014). Speech production after glossectomy: Methodological aspects. *Clinical Linguistics & Phonetics*, **28**(4), 241-256. (doi:10.3109/02699206.2013.802015).
- Adler-Bock, M., Bernhardt, B.M., Gick, B. & Bacsfalvi, P. (2007). The use of ultrasound in remediation of north American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology*, **16**, 128 - 139. 10.1044/1058-0360(2007/017).
- Auzou, P. & Rolland-Monnoury, V. (2006). *BECD: batterie d’évaluation clinique de la dysarthrie*: Ortho éditions.
- Bacsfalvi, P., Bernhardt, B.M. & Gick, B. (2007). Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *Advances in Speech Language Pathology*, **9**(1), 36-45. 10.1080/14417040601101037.
- Badin, P., Tarabalka, Y., Elisei, F. & Bailly, G. (2010). Can you 'read' tongue movements? Evaluation of the contribution of tongue display to speech understanding. *Speech Communication*, **52**(6), 493-503.
- Benoît, C. & Le Goff, B. (1998). Audio-visual speech synthesis from French text: Eight years of models, designs and evaluation at the ICP. *Speech Communication*, **26**, 117-129.
- Bernhardt, B.M., Bacsfalvi, P., Adler-Bock, M., Shimizu, R., Cheney, A., Giesbrecht, N., O'connell, M., Sirianni, J. & Radanov, B. (2008). Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. *Clinical Linguistics & Phonetics*, **22**(2), 149-162. doi:10.1080/02699200701801225.

- Bernhardt, B.M., Gick, B., Bacsfalvi, P. & Ashdown, J. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics*, **17**(3), 199-216. doi:10.1080/0269920031000071451.
- Blyth, K.M., McCabe, P., Madill, C. & Ballard, K.J. (2016). Ultrasound visual feedback in articulation therapy following partial glossectomy. *Journal of Communication Disorders*, **61**, 1-15.
- Byun, T.M., Hitchcock, E.R. & Swartz, M.T. (2014). Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research*, **57**(6), 2166-2130. 10.1044/2014_JSLHR-S-14-0034.
- Cavin, M. (2015). The use of ultrasound biofeedback for improving English /r/. *Working Papers of the Linguistics Circle*, **25**(1), 32-41.
- Chen, Y.-P.P., Johnson, C., Lalbakhsh, P., Caelli, T., Deng, G., Tay, D., Erickson, S., Broadbridge, P., El Refaie, A., Doube, W. & Morris, M.E. (2016). Systematic review of virtual speech therapists for speech disorders. *Computer Speech & Language*, **37**, 98-128. <http://dx.doi.org/10.1016/j.csl.2015.08.005>.
- Cleland, J., McCron, C. & Scobbie, J.M. (2013). Tongue reading: Comparing the interpretation of visual information from inside the mouth, from electropalatographic and ultrasound displays of speech sounds. *Clinical Linguistics & Phonetics*, **27**(4), 299-311. 10.3109/02699206.2012.759626.
- Cleland, J., Scobbie, J.M., Heyde, C., Roxburgh, Z. & Wrench, A.A. (2017). Covert contrast and covert errors in persistent velar fronting. *Clinical Linguistics & Phonetics*, **31**(1), 35-55. 10.1080/02699206.2016.1209788.
- Cleland, J., Scobbie, J.M., Nakai, S. & Wrench, A.A. (2015). Helping children learn non-native articulations: The implications for ultrasound-based clinical intervention. In *18th International Congress of Phonetic Sciences*, vol., pp. Glasgow, Scotland.
- Cleland, J., Scobbie, J.M. & Wrench, A.A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics*, **29**(8-10), 575-597. 10.3109/02699206.2015.1016188.
- Enderby, P.M. (1983). *Frenchay dysarthria assessment*. San Diego, Calif: College-Hill Press.
- Engwall, O. (2008). Can audio-visual instructions help learners improve their articulation? — An ultrasound study of short term changes. In *Proceedings of Interspeech 2008*, vol., pp. 2631-2634. Brisbane, Australia.
- Epstein, M.A. (2005). Ultrasound and the IRB. *Clinical Linguistics & Phonetics*, **19**(6-7), 567-572. doi:10.1080/02699200500113889.
- Erber, N.P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders*, **XL**, 481-492.
- Fabre, D., Hueber, T., Girin, L., Alameda-Pineda, X. & Badin, P. (2017). Automatic animation of an articulatory tongue model from ultrasound images of the vocal tract. *Speech Communication*, **93**, 63-75. <https://doi.org/10.1016/j.specom.2017.08.002>.
- Fagel, S. & Madany, K. (2008). A 3-D virtual head as a tool for speech therapy for children. In *Proceedings of Interspeech 2008*, vol., pp. 2643-2646. Brisbane, Australia.
- Fawcett, S., Bacsfalvi, P. & Bernhardt, B.M. (2008). Ultrasound as visual feedback in speech therapy for/r/with adults with Down Syndrome. *Down Syndrome Quarterly*, **10**(1), 4-12.
- Gallagher, L. (2013). *Critical Review: The effectiveness of ultrasound technology as a visual biofeedback tool on the productive speech intelligibility of adolescents and young adults with a hearing impairment*. University of Western Ontario.
- Ghio, A., Lalain, M., Giusti, L., Pouchoulin, G., Robert, D., Rebourg, M., Fredouille, C., Laaridh, I. & Woisard, V. (2018). Une mesure d'intelligibilité par décodage acoustico-

- phonétique de pseudo-mots dans le cas de parole atypique. In *32^{èmes} Journées d'Etude de la Parole*, vol., pp. 285-293. Aix en Provence, France. DOI: 10.21437/JEP.2018-33.
- Gibbon, F. (2013). *Bibliography of electropalatographic (EPG) studies in English (1957-2013)*: University College Cork, Ireland.
- Haldin, C., Acher, A., Kauffmann, L., Hueber, T., Cousin, E., Badin, P., Perrier, P., Fabre, D., Perennou, D., Detante, O., Jaillard, A., Løevenbruck, H. & Baciú, M. (2017). Speech recovery and language plasticity can be facilitated by Sensori-Motor Fusion training in chronic non-fluent aphasia. A case report study. *Clinical Linguistics & Phonetics*, 1-27. 10.1080/02699206.2017.1402090.
- Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, **50**(3), 346-363. 10.1002/bimj.200810425.
- Hueber, T. (2013). Ultraspeech-player: intuitive visualization of ultrasound articulatory data for speech therapy and pronunciation training. In *Interspeech 2013 (14th Annual Conference of the International Speech Communication Association)*, vol., pp. 752-753. Lyon, France.
- Hueber, T., Chollet, G., Denby, B. & Stone, M. (2008). Acquisition of ultrasound, video and acoustic speech data for a silent-speech interface application. In *8th International Seminar on Speech Production, ISSP8* (R. Sock, S. Fuchs & Y. Laprie, editors), vol., pp. 365-368. Strasbourg, France.
- Katz, W.F. & McNeil, M.R. (2010). Studies of articulatory feedback treatment for apraxia of speech based on electromagnetic articulography. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, **20**(3), 73-79. 10.1044/nnsld20.3.73.
- Massaro, D.W. & Light, J. (2004). Using visible speech to train perception and production of speech for individuals with hearing loss. *Journal of Speech, Language, and Hearing Research*, **47**, 304-320.
- Menin-Sicard, A., Sicard, E. & Bézard, M. (2016). Intérêt de la visualisation de la position et du mouvement des articulateurs pour améliorer l'intelligibilité : la plate-forme Diadolab. In *XVI^{èmes} Rencontres Internationales d'Orthophonie. Orthophonie et technologies innovantes* (N. Joyeux & S. Topouzkhianian, editors), pp. 261-289. Paris, France: Ortho Edition.
- Mental, R., Carey, H., Lee, G.S., Hodge, M.J. & Vick, J. (2017). Measuring progress during practice: Motion analysis throughout visual biofeedback treatment for residual speech sound errors. *The Journal of the Acoustical Society of America*, **142**(4), 2639-2639. 10.1121/1.5014670.
- Mills, A.E. (1987). The development of phonology in the blind child. In *Hearing by eye : the psychology of lipreading* (B. Dodd & R. Campbell, editors), pp. 145-161. London: Lawrence Erlbaum Associates.
- Modha, G., Bernhardt, B.M., Church, R. & Bacsfalvi, P. (2008). Case study using ultrasound to treat /r/. *International Journal of Language & Communication Disorders*, **43**(3), 323-329. 10.1080/13682820701449943.
- Perkell, J.S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, **25**(5), 382-407. <http://dx.doi.org/10.1016/j.jneuroling.2010.02.011>.
- Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., Wilhelms-Tricarico, R. & Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, **28**(3), 233-272. <http://dx.doi.org/10.1006/jpho.2000.0116>.
- Perrier, P., Savariaux, C., Lebeau, J. & Magaña, G. (1999). Speech production after tongue surgery and tongue reconstruction. In *14th International Congress of Phonetic Sciences*, vol., pp. 1805-1808. San Francisco, USA.

- Preston, J.L., Brick, N. & Landi, N. (2013). Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology*, **22**(4), 627-643. 10.1044/1058-0360(2013/12-0139).
- Preston, J.L., Holliman-Lopez, G. & Leece, M.C. (2018). Do participants report any undesired effects in ultrasound speech therapy? *American Journal of Speech-Language Pathology*, **27**(2), 813-818. 10.1044/2017_AJSLP-17-0121.
- Preston, J.L. & Leaman, M. (2014). Ultrasound visual feedback for acquired apraxia of speech: A case report. *Aphasiology*, **28**(3), 278-295. 10.1080/02687038.2013.852901.
- Preston, J.L., Maas, E., Whittle, J., Leece, M.C. & McCabe, P. (2016). Limited acquisition and generalisation of rhotics with ultrasound visual feedback in childhood apraxia. *Clinical Linguistics & Phonetics*, **30**(3-5), 363-381. 10.3109/02699206.2015.1052563.
- Preston, J.L., McAllister, T., Phillips, E., Boyce, S., Tiede, M., Kim, J.S. & Whalen, D.H. (2018). Treatment for residual rhotic errors with high- and low-frequency ultrasound visual feedback: A single-case experimental design. *Journal of Speech, Language, and Hearing Research*, **61**(8), 1875-1892. 10.1044/2018_JSLHR-S-17-0441.
- R Development Core Team. (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, <http://www.R-project.org>.
- Roxburgh, Z., Cleland, J. & Scobbie, J.M. (2016). Multiple phonetically trained-listener comparisons of speech before and after articulatory intervention in two children with repaired submucous cleft palate. *Clinical Linguistics & Phonetics*, **30**(3-5), 398-415. 10.3109/02699206.2015.1135477.
- Roxburgh, Z., Scobbie, J.M. & Cleland, J. (2015). Articulation therapy for children with cleft palate using visual articulatory models and ultrasound biofeedback. In *18th International Congress of Phonetic Sciences*, vol., pp. Glasgow, Scotland.
- Schwartz, J.-L., Basirat, A., Ménard, L. & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, **25**(5), 336-354. <http://dx.doi.org/10.1016/j.jneuroling.2009.12.004>.
- Shawker, T.H. & Sonies, B.C. (1985). Ultrasound biofeedback for speech training: Instrumentation and preliminary results. *Investigative Radiology*, **20**(1), 90-93.
- Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, **19**(6-7), 455-501.
- Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, **26**(2), 212-215.
- Treille, A., Vilain, C., Hueber, T., Lamalle, L. & Sato, M. (2016). Inside speech: Multisensory and modality-specific processing of tongue and lip speech actions. *Journal of Cognitive Neuroscience*, 1-19. 10.1162/jocn_a_01057.
- Turgeon, C., Prémont, A., Trudeau-Fisette, P. & Ménard, L. (2015). Exploring consequences of short- and long-term deafness on speech production: A lip-tube perturbation study. *Clinical Linguistics & Phonetics*, **29**(5), 378-400. doi:10.3109/02699206.2015.1007527.
- Tutz, G. & Hennevoogl, W. (1996). Random effects in ordinal regression models. *Computational Statistics & Data Analysis*, **22**(5), 537-557. [https://doi.org/10.1016/0167-9473\(96\)00004-7](https://doi.org/10.1016/0167-9473(96)00004-7).
- Woisard, V., Espesser, R., Ghio, A. & Duez, D. (2013). From intelligibility to comprehension, which measurement in practice? *Revue de Laryngologie Otologie Rhinologie*, **134**(1), 27-33.